

استخراج ویژگی نظارت شده تصاویر چهره به منظور افزایش دقت شناسایی

مریم ایمانی* و محمدحسن قاسمیان یزدی

دانشکده مهندسی برق و کامپیوتر، دانشگاه تربیت مدرس، تهران، ایران

چکیده

استخراج ویژگی یک گام مهم برای پردازش و تحلیل داده‌های ابربُعدی در مسائل شناسایی الگو است. تصاویر ابرطیفی اخذ شده از سنجنده‌های راه دور و تصاویر چهره انسان از جمله داده‌های ابربُعدی محسوب می‌شوند که با وجود تعداد نمونه آموزشی محدود، کاهش ویژگی یک گام پیش پردازش اساسی برای طبقه‌بندی این گونه داده‌ها محسوب می‌شود. در این مقاله، به بررسی و ارزیابی روش‌های نوین استخراج ویژگی می‌پردازیم که تا کنون برای داده چهره استفاده نشده و در همین اواخر برای کاهش ابعاد تصاویر ابرطیفی سنجش از دور پیشنهاد شده‌اند. در این پژوهش، کارایی هفت روش نوین معرفی شده را برای داده ابرطیفی با چهار روش پرکاربرد استخراج ویژگی مورد ارزیابی و مقایسه قرار خواهیم داد. نتایج آزمایش‌ها بر روی دو داده بانک Yale و ORL، برتری تعدادی از این روش‌های نوین را نسبت به روش‌های استخراج ویژگی LDA، NWFE، MMLDA و LPP نظارت شده، از نظر دقت شناسایی، نشان می‌دهند.

واژگان کلیدی: ابرطیفی، استخراج ویژگی، شناسایی چهره، طبقه‌بندی.

Supervised Feature Extraction of Face Images for Improvement of Recognition Accuracy

Maryam Imani* & Hassan Ghassemian

Faculty of Electrical and Computer Engineering, Tarbiat Modares University, Tehran, Iran

Abstract

Dimensionality reduction methods transform or select a low dimensional feature space to efficiently represent the original high dimensional feature space of data. Feature reduction techniques are an important step in many pattern recognition problems in different fields especially in analyzing of high dimensional data. Hyperspectral images are acquired by remote sensors and human face images are one of the high dimensional data types. Because of limitation in the number of training samples, feature reduction is the important preprocessing step for classification of these types of data. Face recognition is one of the main interesting studies in human computer interaction applications. Face is among the most significant biometric characteristics which are used for identification of individuals. Before face recognition, feature reduction is an important processing step. In this paper, we apply the new feature extraction methods, which have been firstly proposed for feature reduction of hyperspectral imagery remote sensing, on the face databases for the first time. In this research, we compare the performance of seven new feature extraction methods with four state-of-the-art feature extraction methods. The proposed methods are Nonparametric Supervised Feature Extraction (NSFE), Clustering Based Feature Extraction (CBFE), Feature Extraction Using Attraction Points (FEUAP), Cluster Space Linear Discriminant Analysis (CSLDA), Feature Space Discriminant Analysis (FSDA), Feature Extraction using Weighted Training samples (FEWT), and Discriminant Analysis- Principal Component 1 (DA-PC1). The experimental results on two face databases, Yale and ORL, show the better

* Corresponding author

* نویسنده عهده‌دار مکاتبات

performance of some new feature extraction methods, from the recognition accuracy point of view compared to methods such as linear discriminant analysis (LDA), non-parametric weighted feature extraction (NWFE), median-mean line discriminant analysis (MMLDA), and supervised locality preserving projection (LPP).

Keywords: Face recognition, Feature extraction, Classification, Hyperspectral.

۱- مقدمه

استخراج ویژگی و طبقه‌بندی داده‌های ابربعدی از مسائل مهم شناسایی الگو هستند که در سال‌های اخیر بسیار مورد توجه قرار گرفته‌اند [1-5]. از آنجایی که بیشتر روش‌های استخراج ویژگی بر روی بردار عمل می‌کنند، برای استفاده از این روش‌ها بر روی داده‌ی چهره، هر تصویر (ماتریس) چهره، در ابتدا، سطر به سطر یا ستون به ستون تبدیل به یک بردار می‌شود [6-8]. در نتیجه ابعاد بردار ویژگی در داده‌ی چهره بسیار بزرگ خواهد بود و با وجود نمونه‌های آموزشی محدود، منجر به پدیده هیوز [9] خواهد شد.

تحلیل ممیز خطی^۱ (LDA) معروف‌ترین روش استخراج ویژگی است که از بیشینه‌کردن پراکندگی‌های بین طبقه‌ای و کمینه‌کردن پراکندگی‌های درون طبقه‌ای جهت استخراج ویژگی بهره می‌برد [10]؛ اما منفرد شدن ماتریس پراکندگی درون طبقه‌ای با استفاده از مجموعه نمونه آموزشی کوچک، باعث دقت کم این روش در داده‌های ابربعدی می‌شود. استخراج ویژگی وزن دار غیر پارامتریک^۲ (NWFE)، با محاسبه میانگین‌های وزن دار، ماتریس‌های پراکندگی را به شکلی غیر پارامتریک تخمین می‌زند و با تنظیم^۳ ماتریس پراکندگی درون طبقه‌ای با مشکل نمونه‌های آموزشی محدود مقابله می‌کند [11]. تحلیل ممیز خطی میانه-میانگین^۴ (MMLDA) روش استخراج ویژگی دیگری است که در سال‌های اخیر معرفی شده و از خط میانه-میانگین در تخمین ماتریس‌های پراکندگی استفاده می‌کند [12]. هنگامی که در میان نمونه‌های آموزشی، نمونه پرتی وجود داشته باشد، میانگین معیار قابل اطمینانی نخواهد بود و استفاده از هر دو معیار میانه و میانگین نتایج پایورتری با استفاده از MMLDA فراهم می‌آورد. تصویر حفظ محل^۵ (LPP) نیز یک روش استخراج ویژگی است که از اطلاعات جدایی‌پذیری طبقه‌ها بهره‌ای نمی‌برد و در مقابل، ساختار محلی داده را با استفاده از یک گراف مجاورت^۶ حفظ می‌کند [13]. این روش استخراج ویژگی

به دو شیوه نظارت‌شده و بدون نظارت قابل پیاده‌سازی است که فرم نظارت‌شده آن مد نظر ما است. در فرم نظارت‌شده LPP، تنها نمونه‌های هم طبقه برای ساخت گراف در نظر گرفته می‌شوند. به عبارت دیگر، تنها شباهت بین نمونه‌های هم طبقه محاسبه و فاصله بین نمونه‌های غیر هم طبقه، ارزیابی نمی‌شود. وقتی دو نمونه x_i و x_j به یک طبقه تعلق داشته باشند، مقادیر در ماتریس شباهت^۷ برابر یک یا یک تابع گوسی $\exp\left(\frac{-\|x_i - x_j\|^2}{t}\right)$ در نظر گرفته می‌شود و در غیر این صورت، مقداری برابر صفر به آن نسبت داده می‌شود.

تاکنون روش‌های استخراج ویژگی متنوعی برای شناسایی تصاویر چهره معرفی شده‌اند. بیشتر روش‌های پایه استخراج ویژگی مانند تحلیل مؤلفه‌های اصلی (PCA)^۸ [10] یا تحلیل ممیز خطی (LDA) روش‌های بر مبنای بردار هستند. به عبارت دیگر برای اعمال این روش‌ها در مسائل شناسایی چهره، در ابتدا مجبوریم که ماتریس تصویر را سطر به سطر یا ستون به ستون به یک بردار تصویر تبدیل کرده و سپس تبدیل استخراج ویژگی مورد نظر را بر روی بردار تصویر اعمال کنیم. به عنوان یک نتیجه، بردار تصویر در یک فضای بعد بالا (به طور کلی فراتر از تعداد نمونه‌ها) قرار می‌گیرد. در ادبیات یادگیری ماشین از این مسئله به بعد بالا و اندازه نمونه کوچک یاد می‌شود که به کاهش دقت شناسایی می‌انجامد. مسأله مشابه (بعد بالا و تعداد نمونه‌های آموزشی محدود) در حوزه طبقه‌بندی تصاویر ابرطیفی در رشته سنجش از دور نیز به چشم می‌خورد. تصاویر ابرطیفی اخذ شده توسط ماهواره‌های سنجش از دور دارای تعداد زیادی باند طیفی هستند. هدف، تعیین برجسب هر پیکسل از تصویر (طبقه‌بندی آن) است. هر پیکسل دارای بعد بالا، به عبارت دیگر دارای بیش از یکصد باند طیفی (ویژگی) بوده، درحالی که تعداد پیکسل‌های برجسب خورده به عنوان نمونه آموزشی، محدود است. از آنجایی که در سال‌های اخیر، روش‌های نوینی برای غلبه بر مسأله تعداد نمونه آموزشی محدود در حوزه تصاویر ابرطیفی سنجش از دور معرفی شده است، در این مقاله بر آن شدیم که کارایی این روش‌ها را در حوزه شناسایی تصاویر چهره نیز مورد آزمایش قرار دهیم.

⁷ Similarity Matrix

⁸ Principal Component Analysis (PCA)

¹ Linear Discriminant Analysis

² Nonparametric Weighted Feature Extraction

³ Regularization

⁴ Median-Mean and Feature Line Embedding

⁵ Locality Preserving Projection

⁶ Adjacency Graph

چندین خوشه تعلق داشته باشند. به میزان احتمال اختصاص هر کدام از نمونه‌های آموزشی به هر کدام از خوشه‌ها، نمونه‌های آزمایشی آن خوشه به‌همراه آن نمونه آموزشی در تخمین ماتریس‌های پراکندگی مشارکت می‌کنند. روش پیشنهادی پنجم، تحلیل تفکیک‌کننده فضای ویژگی^۵ (FSDA) نام دارد [18]. در این روش برای استخراج ویژگی از داده، دو معیار مختلف را با هم در نظر می‌گیریم. معیار نخست این است که ویژگی‌های انتخاب‌شده تا حد امکان از هم متمایز باشند که برای این امر پراکندگی بین ویژگی‌ها را بیشینه می‌کنیم. معیار دوم افزایش جدایی‌پذیری میان طبقه‌هاست که برای این امر نیز، پراکندگی بین طبقه‌ای را بیشینه و پراکندگی درون طبقه‌ای را کمینه می‌کنیم. روش پیشنهادی ششم، استخراج ویژگی با استفاده از نمونه‌های آموزشی وزن‌دار^۶ (FEWT) است [19]. در این روش به هر کدام از ویژگی‌های نمونه آموزشی وزنی تعلق می‌گیرد. اهمیت نسبی هر کدام از ویژگی‌های نمونه آموزشی در پیش‌گویی طبقه آن نمونه به‌دست آمده و به‌عنوان وزنی برای آن ویژگی از نمونه آموزشی در نظر گرفته می‌شود؛ سپس از این نمونه‌های آموزشی وزن‌دار می‌توان در هر روش استخراج ویژگی دلخواه استفاده کرد. روش پیشنهادی هفتم، تحلیل ممیز- مؤلفه اصلی نخست^۷ (DA-PCI) نامیده شده است [20]. در این روش، ماتریس‌های پراکندگی درون طبقه‌ای و بین طبقه‌ای به شکلی غیرپارامتریک و وزن‌دار طراحی شده‌اند. وزن اختصاص‌یافته به هر جفت از نمونه‌های آموزشی که در تخمین ماتریس‌های پراکندگی مشارکت می‌کنند با استفاده از مؤلفه‌های اصلی نظیر آن‌ها محاسبه می‌شود. هرچه مؤلفه اصلی نخست دو نمونه به هم نزدیک‌تر باشد، وزن بیشتری به آن نمونه‌ها اختصاص داده می‌شود. این وزن‌دهی به حفظ ساختار اصلی داده (در کنار افزایش جدایی‌پذیری طبقه‌ها) کمک خواهد کرد.

روش‌های شناسایی چهره به دو دسته کلی روش‌های آماری و روش‌های هوش مصنوعی (از قبیل سیستم‌های فازی، شبکه‌های عصبی و الگوریتم‌های تکاملی) تقسیم می‌شوند. رویکرد اصلی در این مقاله، بررسی روش‌های آماری شناسایی چهره است و در نتیجه مقایسه‌ای با روش‌های شبکه عصبی صورت نگرفته است. ادامه این مقاله به شکل زیر سازمان‌دهی شده است. در بخش ۲، هفت روش جدیدی را که برای استخراج ویژگی از داده ابرطیفی معرفی شده‌اند، به‌اختصار

روش استخراج ویژگی پیشنهادی نخست، استخراج ویژگی نظارت‌شده غیرپارامتریک^۱ (NSFE) نام دارد [14]. در این روش ماتریس‌های پراکندگی درون طبقه‌ای و غیرطبقه‌ای به‌صورتی غیرپارامتریک تعریف می‌شوند. در تولید ماتریس‌های پراکندگی، تنها نیاز به محاسبه میانگین کلی بوده و نیازی به تخمین میانگین طبقه‌ها نیست؛ در نتیجه، در زمانی که تعداد نمونه‌های آموزشی محدودی در دسترس است و تخمین میانگین طبقه‌ها با دقت خوبی حاصل نمی‌شود، این روش کارایی خوبی خواهد داشت. به‌علاوه، به‌دلیل استفاده از خود نمونه‌های آموزشی به جای میانگین طبقه‌ها در تخمین ماتریس پراکندگی بین طبقه‌ای، رتبه ماتریس و در نتیجه تعداد ویژگی‌های استخراج‌شده محدود به تعداد طبقه‌ها نخواهد بود. روش پیشنهادی دوم، استخراج ویژگی بر مبنای خوشه‌بندی^۲ (CBFE) است [15]. در این روش، در هر ویژگی میانگین c طبقه محاسبه شده و در یک بردار قرار می‌گیرد. به این ترتیب به‌ازای هر ویژگی، یک بردار c بعدی تخصیص می‌دهیم. به‌عبارت دیگر، d بردار c بعدی داریم که d تعداد ویژگی‌های اولیه است. بدیهی است اگر بردار اختصاص‌یافته به دو ویژگی مشابه باشد، این دو ویژگی دارای توانایی یکسانی در تمیز طبقه‌ها خواهند بود. براساس این ایده، یک خوشه‌بندی بر روی این بردارها انجام می‌دهیم. روش پیشنهادی سوم استخراج ویژگی با استفاده از نقاط جذب^۳ (FEUAP) نامیده شده است [16]. در این روش از میان نمونه‌های آموزشی هر طبقه، یک نمونه به‌عنوان نقطه جذب آن طبقه در نظر گرفته می‌شود. با استفاده از یک تابع بهینه‌سازی مناسب، نمونه‌های هر طبقه را به نقطه جذب آن طبقه نزدیک (جذب) و از نقطه جذب سایر طبقه‌ها دور می‌کنیم (دفع). به این ترتیب، با انتخاب مناسب نقاط جذب قادر خواهیم بود که جدایی‌پذیری طبقه‌ها را افزایش دهیم. روش پیشنهادی چهارم، تحلیل ممیز خطی فضای خوشه^۴ (CSLDA) نامیده شده است [17]. در این روش، برای تخمین ماتریس‌های پراکندگی درون طبقه‌ای و بین طبقه‌ای، علاوه بر نمونه‌های آموزشی، از نمونه‌های آزمایشی نیز بهره گرفته می‌شود. در این روش نیازی به دانستن یا پیش‌گویی برجسب نمونه‌های آزمایشی نیست. در این روش یک خوشه‌بندی بر روی مجموعه داده آزمایشی انجام می‌شود. از آن جایی که یک داده ابربعدی ممکن است، چندمده باشد، هر کدام از نمونه‌های آموزشی ممکن است به

¹ Nonparametric Supervised Feature Extraction

² Clustering Based Feature Extraction

³ Feature Extraction Using Attraction Points

⁴ Cluster Space Linear Discriminant Analysis

⁵ Feature Space Discriminant Analysis

⁶ Feature Extraction using Weighted Training samples

⁷ Discriminant Analysis- Principal Component 1

مرور خواهیم کرد و در بخش سوم، این روش‌ها را بر روی داده‌های چهره آزمایش و مورد ارزیابی قرار می‌دهیم. در نهایت، جمع‌بندی مقاله در بخش چهارم بیان شده است.

۲- روش‌های نوین استخراج ویژگی

در این بخش به معرفی مختصر روش‌های استخراج ویژگی‌ای می‌پردازیم که برای داده ابرطیفی سنجش از دور پیشنهاد شده و سپس در بخش بعدی به ارزیابی این روش‌ها بر روی داده چهره خواهیم پرداخت.

۲-۱- استخراج ویژگی نظارت‌شده غیر پارامتریک (NSFE)

در این روش، که سعی در فشردن نمونه‌های هم‌طبقه و جداکردن نمونه‌های غیر هم‌طبقه دارد، ماتریس‌های پراکندگی بین-طبقه‌ای (S_b) و درون-طبقه‌ای (S_w) به شکل وزن دار و غیر پارامتریک زیر تعریف شده‌اند:

$$S_b = \sum_{k=1}^c \sum_{i=1}^{n_k} n_k d(x_{i,k}, m)(x_{i,k} - m)(x_{i,k} - m)^T \quad (1)$$

$$S_w = \sum_{k=1}^c \sum_{i=1}^{n_k} \sum_{j=1}^{n_k} d(x_{i,k}, x_{j,k})(x_{i,k} - x_{j,k})(x_{i,k} - x_{j,k})^T \quad (2)$$

که در روابط بالا $d(a, b) = \|a - b\|_2$ به معنای فاصله اقلیدسی بین بردارهای a و b است، $m = \frac{1}{n} \sum_{k=1}^c \sum_{i=1}^{n_k} x_{i,k}$ میانگین کل نمونه‌های آموزشی، $x_{i,k}$ نمونه‌ی i ام در طبقه k ام، n_k تعداد نمونه‌های آموزشی طبقه k ام، c تعداد طبقه‌ها و n تعداد کل نمونه‌های آموزشی است. به دلیل محدود بودن رتبه ماتریس پراکندگی بین طبقه‌ای، تعداد ویژگی‌های استخراج شده توسط این روش، محدود به تعداد طبقه‌ها نیست؛ به علاوه، برای به دست آوردن ماتریس تبدیل استخراج ویژگی $W_{d \times n} = (w_1, w_2, \dots, w_n)$ که از بردار $x_i \in \mathbb{R}^d$ در فضای اولیه با بعد بالا، داده را به $z_i \in \mathbb{R}^n$ در فضای ثانویه با بعد کمتر می‌رساند ($z_i = W^T x_i$) به جای بیشینه کردن معیار $\text{tr}(S_w^{-1} S_b)$ از بیشینه کردن معیار $\text{tr}(S_b - S_w)$ استفاده شده است؛ در نتیجه، به دلیل نیاز نداشتن به محاسبه معکوس ماتریس پراکندگی درون طبقه‌ای، با مشکل منفرد شدن S_w در نمونه‌های آموزشی محدود نیز، برخورد نخواهیم کرد.

۲-۲- استخراج ویژگی بر مبنای خوشه‌بندی (CBFE)

روش استخراج ویژگی نظارت‌شده بر مبنای خوشه‌بندی (CBFE) تنها از آمارگان مرتبه نخست (بردار میانگین) استفاده می‌کند. در نتیجه در شرایط بدرفتار و با استفاده از

نمونه آموزشی محدود، کارایی خوبی دارد. در این روش، ابتدا بردار میانگین نمونه‌های آموزشی در هر طبقه را محاسبه و مقادیر به دست آمده را در یک ماتریس A قرار می‌دهیم.

$$A = \begin{bmatrix} m_{11} & m_{12} & \dots & m_{1c} \\ m_{21} & m_{22} & \dots & m_{2c} \\ \vdots & \vdots & \ddots & \vdots \\ m_{d1} & m_{d2} & \dots & m_{dc} \end{bmatrix} \quad (3)$$

در رابطه بالا m_{ij} میانگین نمونه‌های آموزش طبقه j در بعد i ام است. ماتریس A با استفاده از بردارهای سطری a_i ($i = 1, 2, \dots, d$) می‌تواند به شکل زیر بازنویسی شود:

$$A = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_d \end{bmatrix} \quad (4)$$

هر بردار a_i شامل مقادیر میانگین نمونه‌های آموزشی برای c طبقه است. شباهت دو بردار a_i و a_j ($i \neq j$) به یکدیگر، به طور ضمنی بیان می‌کند که ویژگی‌های i و j به مقدار زیاد همبسته بوده و یکی از آن‌ها می‌تواند حذف شود. بر مبنای این ایده، ما با استفاده از یک الگوریتم خوشه‌بندی مثل k -means بردارهای a_i ($i = 1, 2, \dots, d$) را به چندین خوشه گروه‌بندی می‌کنیم. بعد از خوشه‌بندی، میانگین ویژگی‌هایی که بردار a_i نظیر آن‌ها در یک خوشه قرار گرفته‌اند، به عنوان یک ویژگی جدید استخراج شده در نظر گرفته می‌شوند. در روش CBFE، تعداد خوشه‌های انتخابی برابر تعداد ویژگی‌های استخراج شده است و اگر d تعداد ویژگی‌های اولیه داده باشد، داریم: $1 \leq k \leq d$. روش CBFE بر خلاف روش‌هایی مانند LDA که تنها توانایی استخراج حداکثر $c - 1$ ویژگی را دارند، قادر به استخراج هر تعداد ویژگی دلخواه است.

۲-۳- استخراج ویژگی با استفاده از نقاط جذب (FEUAP)

تصویر ابرطیفی را در فضای ویژگی با بعد بالا با $\{x_i\}_{i=1}^N, x_i \in \mathbb{R}^d$ و تبدیل یافته آن در فضای ویژگی با بعد کاهش یافته را با $\{y_i\}_{i=1}^N, y_i \in \mathbb{R}^m$ نشان می‌دهیم. d تعداد باندهای طیفی (ویژگی‌های) اولیه و m تعداد ویژگی‌های استخراج شده در زیر فضای کاهش یافته است ($m \leq d$). ماتریس تبدیل $A_{m \times d}$ هر نقطه از داده اولیه (x_i) را به فضای ثانویه (y_i) نگاشت می‌کند: $y_i = Ax_i$. هدف ما استخراج ویژگی داده ابرطیفی است؛ به گونه‌ای که داده در فضای کاهش یافته نیز برای طبقه‌بندی مناسب باشد. در نتیجه، باید تبدیل را به گونه‌ای انجام دهیم که در فضای تبدیل یافته (کاهش یافته) تمایز بین

$$F(x_{ic}) = \sum_{j=1}^{n_{tc}} e^{-\|x_{ic} - x_{jc}\|^2} \quad (6)$$

سیس، برای به‌دست‌آوردن نقطه جذب طبقه c م مسئله بهینه‌سازی زیر را حل می‌کنیم:

$$s = \arg \max_{i=1, \dots, n_{tc}} F(x_{ic}) \quad (7)$$

در نتیجه، نمونه s ام از طبقه c ام به‌عنوان نقطه جذب طبقه c انتخاب می‌شود، به‌عبارت دیگر خواهیم داشت: $x^{ac} = x_{sc}$. برای مشخص‌کردن تابع تبدیل مناسب برای تولید فضای ویژگی با بعد کاهش‌یافته، باید توجه داشته باشیم که نمونه‌ها در فضای کاهش‌یافته، باید به‌گونه‌ای باشند که:

۱- هر نمونه، تا نقطه جذب طبقه خودش، کمترین فاصله را داشته باشد (جذب).

۲- هر نمونه، تا نقاط جذب سایر طبقه‌ها، بیشترین فاصله را داشته باشد (دفع).

بر مبنای ایده بیان‌شده، دو تابع جذب (ψ_1) و دفع (ψ_2) را به‌شکل زیر تعریف می‌کنیم:

$$\psi_1 = \sum_{c=1}^{n_c} \sum_{i=1}^{n_{tc}} \|y_{ic} - y^{ac}\|^2 \quad (8)$$

$$\psi_2 = - \sum_{c=1}^{n_c} \sum_{i=1}^{n_{tc}} \sum_{\substack{k=1 \\ k \neq c}}^{n_{tc}} \|y_{ic} - y^{ak}\|^2 \quad (9)$$

$y^{ac} = Ax^{ac}$ نقطه جذب طبقه c ام در فضای ویژگی جدید است. هم‌چنین داریم: $y_{ic} = Ax_{ic}$. برای به‌دست‌آوردن ماتریس تبدیل A ، باید مسئله بهینه‌سازی را حل کنیم:

$$\min_A (\psi = \psi_1 + \psi_2) \quad (10)$$

تابع ψ می‌تواند به‌شکل دیگری نیز نوشته شود:

$$\psi = \sum_{j=1}^n \sum_{i=1}^n w_{ij} \|y_i - y_j\|^2 \quad (11)$$

که:

$$w_{ij} = \begin{cases} 1 & \text{if } C1 \\ -1 & \text{if } C2 \\ 0 & \text{if } C3 \end{cases}$$

شروط $C1$ ، $C2$ و $C3$ به شرح زیر هستند:

$C1$: y_i و y_j هم‌طبقه باشند و یکی از آن‌ها نقطه جذب باشد.

$C2$: y_i و y_j غیر هم‌طبقه باشند و دست‌کم یکی از آن‌ها نقطه جذب باشد.

$C3$: هیچ‌کدام از y_i یا y_j نقطه جذب نباشند.

در رابطه (۱۱)، n تعداد کل نمونه‌های آموزشی است. $(n = \sum_{c=1}^{n_c} n_{tc})$. $y_i = Ax_i$ و x_i نمونه آموزشی i ام است.

تابع بهینه‌سازی در (۱۱) می‌تواند به‌شکل ماتریسی زیر بازنویسی شود:

$$\psi = 2tr(YGY^T) \quad (12)$$

طبقه‌ها افزایش یافته و در فضای ویژگی جدید، طبقه‌ها قابل جداسازی از هم باشند. روش پیشنهادی بر اساس ایده زیر شکل گرفته است [16]: می‌توان برای هر طبقه، یک نقطه جذب (مرجع) در نظر گرفت؛ به‌گونه‌ای که نقاط جذب طبقه‌های مختلف در فواصل دور از هم قرار گرفته باشند؛ سیس باید تبدیلی انجام دهیم که نقاط آموزشی هر طبقه به سمت نقطه جذب خود حرکت کرده، به سمت آن گرایش پیدا کنند و در اطراف آن جمع شوند. به این ترتیب تمایز بین طبقه‌ها در فضای جدید رخ می‌دهد (به شکل (۱) توجه کنید). برای انجام روش پیشنهادی، باید دو هدف زیر را دنبال کنیم:

۱- تعیین نقاط جذب هر طبقه

۲- تعیین تبدیلی که فرآیند جذب و کشش نقاط هر طبقه را به سمت نقطه جذب متناظر آن طبقه، انجام دهد.

نقاط جذب از میان نمونه‌های آموزشی هر طبقه انتخاب می‌شوند. ما در این قسمت دو روش برای تعیین نقاط جذب طبقه‌ها پیشنهاد می‌دهیم. روش نخست بر مبنای فاصله و روش دوم بر مبنای معیار تراکم است که به‌ترتیب در ادامه بیان می‌شوند:

تعیین نقاط جذب با استفاده از معیار فاصله: در این روش، نقطه جذب هر کلاس از میان نمونه‌های آموزشی همان طبقه انتخاب می‌شود، به‌گونه‌ای که:

۱- نقطه جذب هر طبقه کمترین فاصله را از نمونه‌های هم‌طبقه خود دارد.

۲- نقطه جذب هر طبقه، بیشترین فاصله را از نمونه‌های غیر هم‌طبقه خود دارد.

به‌عبارت دیگر، برای تعیین نقطه جذب طبقه c ام (x^{ac})، مسئله بهینه‌سازی زیر باید حل شود:

$$s = \arg \min_{i=1, \dots, n_{tc}} \left(\sum_{j=1}^{n_{tc}} \|x_{ic} - x_{jc}\|^2 - \sum_{\substack{k=1 \\ k \neq c}}^{n_c} \sum_{j=1}^{n_{tc}} \|x_{ic} - x_{jk}\|^2 \right) \quad (13)$$

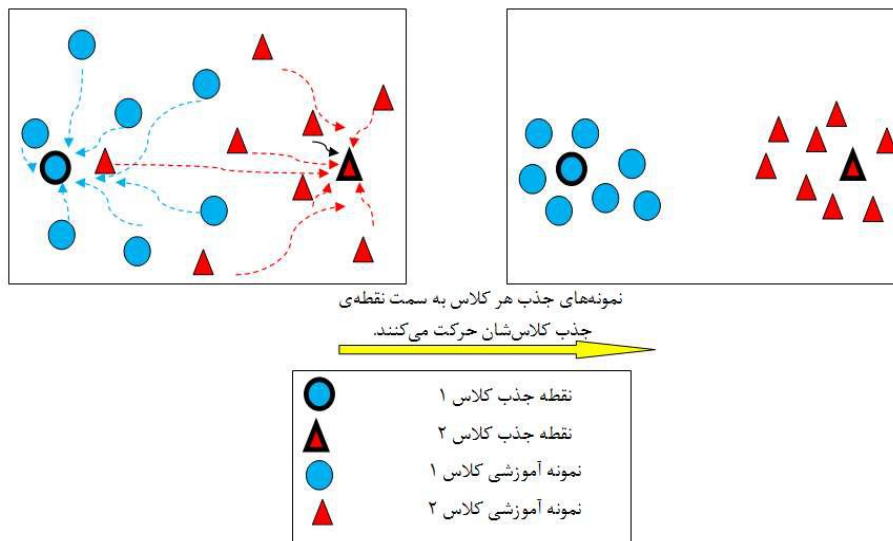
که در نتیجه داریم: $x^{ac} = x_{sc}$ در رابطه بالا، x_{ic} نمونه s ام از طبقه c ام، n_{tc} تعداد نمونه‌های آموزشی طبقه c ام و n_c تعداد طبقه‌ها است.

تعیین نقاط جذب با استفاده از معیار تراکم: در این روش، از میان نمونه‌های آموزشی هر طبقه، نمونه‌ای به‌عنوان نقطه جذب انتخاب می‌شود که تعداد بیشتری از نمونه‌های آن طبقه حول آن تجمع داشته باشند. به‌عبارت دیگر، نقطه جذب در این روش، نقطه‌ای است که تراکم اصلی طبقه، در اطراف آن نقطه شکل گرفته باشد. در این روش، برای هر نمونه یک تابع تراکم به‌شکل زیر تعریف می‌کنیم:

که Y یک ماتریس $d \times n$ بوده، $Y = [y_1, y_2, \dots, y_n]$ و $G = D - W$ ، همچنین، AX ، ($X = [x_1, x_2, \dots, x_n]$) که w_{ij} ($i = 1:n, j = 1:n$) عناصر سازنده ماتریس W بوده، D یک ماتریس قطری است و $D_{ii} = \sum_j w_{ij}$. در نتیجه مسئله بهینه‌سازی (۱۲)، به شکل زیر قابل بازنویسی است:

$$\min_A \psi = 2tr(AXGX^T A^T) \quad (13)$$

ماتریس تبدیل خطی بهینه A از حل مسئله بهینه‌سازی مقدار ویژه بالا به دست می‌آید. برای تبدیل داده از فضای ویژگی با بعد بالا به فضای ویژگی با بعد کاهش یافته m ، به گونه‌ای که $A^T A = I_{m \times m}$ ، بردار ویژه ماتریس XGX^T متناظر با m تا از کوچک‌ترین مقادیر ویژه، تابع تبدیل A را می‌سازند.



(شکل-۱): حرکت به سمت نقاط جذب در یک فضای ویژگی دو بعدی برای یک داده دو کلاسه (Figure-1): moving toward attraction points in a two-dimensional feature space for a two-class dataset.

$$M = \left(\frac{1}{n_c}\right) \sum_{c=1}^{n_c} m_c + \left(\frac{1}{K}\right) \sum_{k=1}^K \mu_k \quad (17)$$

در روابط بالا، x_{ic} نمونه آموزشی نام از طبقه c ام، n_{tc} تعداد نمونه‌های آموزشی طبقه c ام، x^k یک نمونه آزمایشی متعلق به خوشه k ام و X^k مجموعه‌ای از نمونه‌های آزمایشی است که به خوشه k ام تعلق دارند. K تعداد خوشه‌ها و $p(k|x_{ic})$ احتمال پسین تعلق نمونه آموزشی i ام از طبقه c ام (x_{ic}) به خوشه k ام است. M_c میانگین ترکیبی طبقه c و M میانگین ترکیبی کل است. همچنین، $m_c = \frac{1}{n_{tc}} \sum_{i=1}^{n_{tc}} x_{ic}$ میانگین نمونه‌های آموزشی طبقه c ام و $\mu_k = \frac{1}{N_k} \sum_{x^k \in X^k} x^k$ میانگین نمونه‌های آزمایشی خوشه k ام، هستند. در این روش همانند روش NWFE از تنظیم ماتریس S_w برای جلوگیری از منفرد شدن آن و افزایش دقت، استفاده می‌کنیم. در انتها، پس از محاسبه ماتریس‌های پراکندگی، با بیشینه‌کردن معیار $tr(S_w^{-1} S_b)$ ، ماتریس تبدیل استخراج ویژگی به دست خواهد آمد. از روش خوشه‌بندی پرکاربرد k -means با $K = n_c$ به منظور خوشه‌بندی در [17] استفاده شده و تعداد خوشه‌ها برابر تعداد طبقه‌ها ($K = n_c$) در نظر گرفته شده است. یک ایراد روش CSLDA این است که رتبه ماتریس S_b

۲-۴- تحلیل ممیز خطی فضای خوشه (CSLDA)

روش استخراج ویژگی CSLDA بدون نیاز به تعیین برجسته نمونه‌های آزمایشی، با انجام خوشه‌بندی بر روی آن‌ها و یافتن رابطه بین نمونه‌های آموزشی و آزمایشی، از قدرت نمونه‌های بدون برجسته به همراه نمونه‌های آموزشی، برای تولید ماتریس‌های پراکندگی درون طبقه‌ای و بین طبقه‌ای استفاده می‌کند. در این روش، بر روی داده آزمون، یک خوشه‌بندی با K خوشه انجام می‌دهیم. حال از خوشه‌بندی به دست آمده، ارتباط بین نمونه‌های آموزشی و آزمایشی را به دست آورده و از آن‌ها برای تخمین ماتریس‌های پراکندگی درون طبقه‌ای (S_w) و بین طبقه‌ای (S_b) استفاده می‌کنیم:

$$S_w = \sum_{c=1}^{n_c} \sum_{i=1}^{n_{tc}} [(x_{ic} - m_c)(x_{ic} - m_c)^T + \sum_{k=1}^K \sum_{x^k \in X^k} p(k|x_{ic}) (x^k - \mu_k)(x^k - \mu_k)^T] \quad (14)$$

$$S_b = \sum_{c=1}^{n_c} n_{tc} (M_c - M)(M_c - M)^T \quad (15)$$

$$M_c = m_c + \left(\frac{1}{n_{tc}}\right) \sum_{i=1}^{n_{tc}} \sum_{k=1}^K p(k|x_{ic}) \mu_k \quad (16)$$

که

در روش LPP نظارت‌شده از نمونه‌های آموزشی وزن‌دار استفاده شده است.

فرض کنید $X_{n \times d} = [x_1, x_2, \dots, x_n]^T$ مجموعه نمونه‌های آموزشی باشد که $n = \sum_{k=1}^c n_k$ تعداد کل نمونه‌های آموزشی، n_k تعداد نمونه آموزشی طبقه k ، c تعداد طبقه‌ها و d تعداد ویژگی‌ها (باندای طیفی) است. نمونه آموزشی i μ یک بردار d بعدی به صورت $x_i = [x_{i1}, x_{i2}, \dots, x_{id}]$ ؛ $1 \leq i \leq n$ است که ویژگی j μ از نمونه i ام است. فرض کنید $L_{n \times c}$ ماتریس برچسب طبقه‌ها باشد؛ ماتریس L حاوی مقادیر صفر و یک است. اگر نمونه i ام متعلق به طبقه k باشد، در سطر i ام تنها ستون k ام مقدار یک و سایر ستون‌ها مقدار صفر دارند. با توجه به اهمیت نسبی هر ویژگی در پیش‌گویی طبقه هر یک از نمونه‌های آموزشی، می‌توان رابطه بین داده آموزشی و برچسب طبقه‌ها را به صورت زیر مدل کرد:

$$L_{n \times c} = X_{n \times d} W_{d \times c} + 1_n b^T \quad (18)$$

که 1_n یک بردار ثابت $n \times 1$ بعدی است که همه المان‌های آن یک هستند و $b \in \mathbb{R}^{c \times 1}$ عبارت بایاس است. W ماتریس وزن است و w_{jk} ($j = 1, \dots, d$; $k = 1, \dots, c$) المان‌های آن را تشکیل می‌دهند. w_{jk} اهمیت نسبی ویژگی j ام در پیش‌گویی طبقه k ام را بیان می‌کند. برای به‌دست آوردن ماتریس وزن W از روش ساده و کارای LS به صورت زیر استفاده می‌کنیم:

$$\min_{W, b} \psi(W, b) = \|XW + 1_n b^T - L\|_F^2 + \gamma \|W\|_F^2 \quad (19)$$

که $\|\cdot\|_F$ بیان‌گر نرم فروبنیوس است. جمله دوم عبارت بالا، به منظور تنظیم استفاده شده و $\gamma > 0$ یک پارامتر مصالحه است. ماتریس W و بردار b به آسانی بر مبنای تئوری ماتریس به دست می‌آیند [21]:

$$W = (X^T H X + \gamma I_d)^{-1} X^T H L \quad (20)$$

$$b = \frac{L^T 1_n - W^T X^T 1_n}{n} \quad (21)$$

که در رابطه بالا: $H = I_n - (1/n) 1_n 1_n^T$ بوده و I_n بیان‌گر ماتریس همانی $n \times n$ است. پس از محاسبه ماتریس W از آن برای وزن‌دهی نمونه‌های آموزشی استفاده می‌کنیم. اگر نمونه آموزشی i ام (x_i) به طبقه k تعلق داشته باشد، $x_i \in \pi_k$ ، در آن صورت، نحوه وزن‌دهی به صورت زیر خواهد بود:

$$z_{ij} = x_{ij} w_{jk}; 1 \leq j \leq d \quad (22)$$

محاسبه‌شده، برابر تعداد طبقه‌ها (n_c) است که سبب می‌شود، حداکثر بتوان n_c ویژگی با استفاده از این روش، استخراج کرد.

۵-۲- تحلیل تفکیک‌کننده فضای ویژگی (FSDA)

در روش تحلیل تفکیک‌کننده فضای ویژگی FSDA برای استخراج ویژگی‌هایی که قدرت تمیز بین طبقه‌ها را افزایش داده و دقت طبقه‌بندی را افزایش دهند، دو معیار را با هم در نظر می‌گیریم. ویژگی‌های استخراج‌شده باید به گونه‌ای باشند که: ۱- ویژگی‌های تولیدشده، تا حد ممکن از هم متفاوت باشند. ۲- جدایی‌پذیری بین طبقه‌ها افزایش یابد. برای به‌دست آوردن تبدیل استخراج ویژگی FSDA، در مرحله نخست باید ویژگی‌ها را تا حد ممکن، متفاوت از هم استخراج و در مرحله بعدی جدایی‌پذیری طبقه‌ها را بیشینه کنیم. به این منظور، در ابتدا، ماتریس پراکندگی بین ویژگی را به دست آورده و با بیشینه کردن آن، فضای ویژگی را به فضای ویژگی دیگری تبدیل می‌کنیم، به گونه‌ای که پراکندگی بین ویژگی‌ها در آن فضا افزایش یابد (ویژگی‌ها در آن فضا تا جای ممکن از هم متفاوت باشند) و سپس برای افزایش تفکیک‌پذیری بین طبقه‌ها، هم‌زمان، پراکندگی‌های بین طبقه‌ای را بیشینه و پراکندگی‌های درون طبقه‌ای را کمینه می‌کنیم.

۶-۲- استخراج ویژگی با استفاده از نمونه‌های آموزشی وزن دار (FEWT)

در تعیین طبقه نمونه‌ها همه ویژگی‌ها نقش یکسانی ندارند. بلکه بعضی از ویژگی‌ها نقش عمده‌تری در شناسایی طبقه‌ها ایفا می‌کنند. با این حال، روش‌های معمول استخراج ویژگی، این اهمیت نسبی ویژگی‌ها در تعیین طبقه نمونه‌ها را در نظر نمی‌گیرند و وزن یکسانی برای تمام ویژگی‌های یک نمونه آموزشی قائل هستند. در روش FEWT، اهمیت نسبی هر کدام از ویژگی‌های نمونه آموزشی در پیش‌گویی طبقه آن نمونه، به صورت وزنی برای آن ویژگی در آن نمونه آموزشی در نظر گرفته می‌شود. به عبارت دیگر، به تک‌تک ویژگی‌های هر نمونه آموزشی، وزنی نسبت داده می‌شود. پس از وزن‌دار کردن نمونه‌های آموزشی، مجموعه نمونه‌های آموزشی وزن‌دار برای استخراج ویژگی استفاده می‌شود. توجه کنید که هر روش استخراج ویژگی نظارت‌شده، در این مرحله می‌تواند مورد استفاده قرار گیرد که در [19] از روش LPP نظارت‌شده استفاده شده است. به عبارت دیگر برای تولید ماتریس شباهت

که x_{ij} ویژگی j ام از نمونه آموزشی i ام، w_{jk} وزن ویژگی j ام در طبقه k ام و z_{ij} ویژگی j ام از نمونه وزن داده شده i ام است. به این ترتیب، نمونه‌های آموزشی $(1 \leq i \leq n)$ به نمونه‌های آموزشی وزن دار $(1 \leq i \leq n)$ تبدیل می‌شوند. حال از مجموعه نمونه‌های آموزشی وزن دار برای استخراج ویژگی استفاده می‌کنیم. توجه کنید که هر روش استخراج ویژگی، در این مرحله می‌تواند مورد استفاده قرار گیرد که ما در این جا از روش LPP نظارت شده استفاده کرده‌ایم. به عبارت دیگر برای تولید ماتریس شباهت در روش LPP نظارت شده از نمونه‌های آموزشی وزن دار استفاده کرده‌ایم. با فرض این که $y_i = Az_i$ تبدیل یافته z_i در فضای کاهش یافته باشد، ماتریس تبدیل A در روش LPP از حل مسأله بهینه‌سازی زیر به دست می‌آید:

$$\min (\Phi = \sum_{i=1}^n \sum_{j=1}^n \|y_i - y_j\|^2 s_{ij}) \quad (23)$$

که

$$s_{ij} = \begin{cases} 1 & z_i, z_j \text{ belong to a class} \\ 0 & z_i, z_j \text{ do not belong to a class} \end{cases}$$

اگر برای داده با بُعد بالا در فضای اولیه و داده با بُعد کاهش یافته در فضای ثانویه، به ترتیب فرم‌های ماتریسی Z و Y را به صورت زیر در نظر بگیریم:

$$Z = [z_1, z_2, \dots, z_n]_{d \times n}$$

$$Y = [y_1, y_2, \dots, y_n]_{d \times n}$$

می‌توان تابع Φ را به شکل ماتریسی بازنویسی کرد:

$$\Phi = \text{tr}(YLY^T) \quad (24)$$

$$\min_A [\Phi = \text{tr}(AZLZ^T A^T)] \quad (25)$$

که در روابط بالا، $Y = AZ$ و $L = D - S$ است. S ماتریس شباهت است که عناصر آن را $s_{ij} (i = 1:n, j = 1:n)$ تشکیل می‌دهند و D یک ماتریس قطری است که المان‌های روی قطر آن، مجموع سطری ماتریس S است. برای استخراج m ویژگی از d ویژگی اولیه ($m < d$)، تنها کافی است که m بردار ویژه متناظر با m تا کوچک‌ترین مقادیر ویژه ماتریس ZLZ^T داخل ماتریس A بریزیم. در آن صورت، هر نمونه در فضای کاهش یافته، یک بردار m بُعدی به شکل $y_{m \times 1} = A_{m \times d} z_{d \times 1}$ خواهد بود.

۷-۲- تحلیل ممیز - مؤلفه اصلی نخست (DA-PC1)

روش استخراج ویژگی DA-PC1 از تحلیل ممیز برای افزایش جدایی پذیری طبقه‌ها استفاده می‌کند. در این روش، پراکندگی‌های درون طبقه‌ای کمینه و پراکندگی‌های بین طبقه‌ای بیشینه می‌شود. ماتریس‌های پراکندگی به شکل غیرپارامتریک و وزن دار طراحی شده‌اند، به گونه‌ای که مشکل محدودیت تعداد ویژگی‌های استخراجی و منفرد شدن ماتریس پراکندگی درون طبقه‌ای از بین برود. در عین حال برای وزن دهی ماتریس‌های پراکندگی از مؤلفه اصلی نخست (PC1) استفاده شده که این خود سبب حفظ ساختار غالب داده کاهش یافته پس از تبدیل خواهد شد. ماتریس‌های پراکندگی درون طبقه‌ای و بین طبقه‌ای در روش پیشنهادی DA-PC1 به صورت زیر محاسبه می‌شوند:

$$S_w = \sum_{j=1}^n \sum_{i=1}^n w_{ij} (x_{ti} - x_{tj})(x_{ti} - x_{tj})^T \quad (26)$$

$$S_b = \sum_{j=1}^n \sum_{i=1}^n w_{ij} (x_{ti} - x_{tj})(x_{ti} - x_{tj})^T \quad (27)$$

که در روابط بالا، x_{ti} نمونه آموزشی i ام و n تعداد کل نمونه‌های آموزشی است. $l_i \in \{1, 2, \dots, c\}$ برچسب نمونه x_{ti} بوده و c تعداد طبقه‌ها است. وقتی مؤلفه‌های اصلی دو نمونه، به هم نزدیک‌ترند، با وزن بیشتری در تخمین ماتریس‌های پراکندگی مشارکت خواهند داشت. در نتیجه، وزن w_{ij} از رابطه زیر به دست می‌آید:

$$w_{ij} = \frac{1}{|PC1(x_{ti}) - PC1(x_{tj})|^2 + 1} \quad (28)$$

عدد یک که به مخرج اضافه شده، برای جلوگیری از بی‌نهایت شدن مقدار w_{ij} است. در رابطه بالا داریم:

$$PC1(x_{ti}) = v_1^T x_{ti} \quad (29)$$

$PC1(x_{ti})$ مؤلفه اصلی نخست برای هر نمونه از داده است که با استفاده از بردار ویژه نظیر بزرگترین مقدار ویژه ماتریس کواریانس $\Sigma_x = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(x_i - \bar{x})^T$ یعنی v_1 به دست می‌آید. همچنین، $x_i \in \mathcal{R}^d (i = 1, 2, \dots, N)$ نمونه i ام داده، d تعداد ویژگی‌های اولیه، N تعداد کل نمونه‌های داده و $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$ میانگین کل نمونه‌های داده است. به این ترتیب، روش DA-PC1 علاوه بر اطلاعات جدایی پذیری طبقه‌ها از اطلاعات موجود در مؤلفه اصلی نخست (PC1) نیز استفاده می‌کند. مؤلفه PC1 بزرگترین واریانس و قدرت داده

را دارد و در نتیجه ساختار غالب داده پس از تبدیل و کاهش بُعد با استفاده از آن حفظ می‌شود.

در بخش بعد، به آزمایش و ارزیابی روش‌های استخراج ویژگی بالا بر روی داده‌های چهره می‌پردازیم.

۳- اعمال روش‌های پیشنهادی استخراج ویژگی بر روی داده چهره

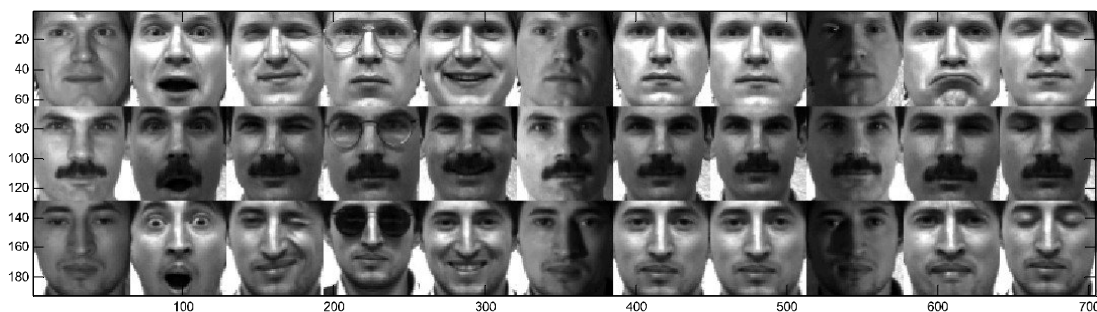
در این قسمت به بررسی روش‌های استخراج پیشنهادی خود برای داده‌های ابرطیفی بر روی داده‌های چهره می‌پردازیم. آزمایش‌ها بر روی یک لپ‌تاپ با سیستم عامل ۶۴ بیتی، ویندوز ۷ با 4G حافظه و پردازش‌گر Intel Core i5-520 M Processor 2.40 GHz انجام شده‌اند. جهت پیاده‌سازی روش‌های شناسایی چهره از نرم‌افزار MATLAB نسخه R2014a بهره گرفته شده است. توجه به این نکته ضروری است که ماهیت تصاویر ابرطیفی سنجش از دور با تصاویر چهره متفاوت است. تصاویر ابرطیفی، به صورت یک مکعب هستند؛ به گونه‌ای که متناظر با هر پیکسل از تصویر یک بردار وجود دارد (که حاوی مقدار آن پیکسل در باندهای طیفی مختلف است)، در حالی که تصاویر چهره به صورت یک ماتریس دوبعدی هستند؛ جایی که به ازای هر پیکسل از تصویر یک مقدار اسکالر وجود دارد. در طبقه‌بندی تصاویر ابرطیفی، هدف برچسب‌زدن به هر پیکسل از تصویر است که خود پیکسل با یک بردار ویژگی مشخص می‌شود؛ در حالی که در شناسایی تصاویر چهره، هدف برچسب‌زدن به کل یک تصویر است؛ جایی که با تغییر شکل ماتریس تصویر به یک بردار، کل تصویر با یک بردار ویژگی نمایش داده می‌شود. در تصاویر ابرطیفی نمونه‌های مختلف از طبقه‌های مختلف درون مکعب تصویر واقع‌اند؛ در حالی که در تصاویر چهره، نمونه‌های مختلف مربوط به طبقه‌ها (افراد) مختلف، تصاویر مجزایی داخل دیتاست هستند. توجه به نکات یادشده جهت پیاده‌سازی و اعمال روش‌های استخراج ویژگی پیشنهادشده برای تصاویر ابرطیفی بر روی تصاویر چهره لازم و ضروری است که در این جا رعایت شده است. برای انجام شناسایی چهره، ابتدا زیرفضای ویژگی با استفاده از روش استخراج ویژگی (کاهش ویژگی) به دست آمده و داده‌های چهره به داخل آن زیرفضا تصویر می‌شوند و سپس از طبقه‌بند نزدیکترین همسایگی (NN) با فاصله اقلیدسی برای شناسایی چهره جدید (نمونه آزمایش) استفاده

می‌شود. NN یک طبقه‌بند ساده و غیرپارامتریک است که دارای دقت و کارایی خوبی برای طبقه‌بندی تصاویر چهره است. این طبقه‌بند نیاز به هیچ تنظیمات خاصی ندارد و ساده و سریع اجرا می‌شود. از آن جایی که هدف ما در این مقاله، مقایسه روش‌های استخراج ویژگی تصاویر چهره است، جهت مقایسه منصفانه بین همه روش‌ها، ویژگی‌های استخراج‌شده توسط همه روش‌های مورد مطالعه را با استفاده از این طبقه‌بند، دسته‌بندی کرده و شناسایی تصاویر چهره را انجام می‌دهیم. جهت مقایسه منصفانه لازم است که از نمونه‌های آموزشی یکسان و از طبقه‌بند یکسان با پارامترهای یکسان استفاده شود که این امر در مقاله رعایت شده است. دقت شناسایی با استفاده از رابطه زیر محاسبه شده است:

$$\text{Recognition Accuracy} = (\text{num_correct} / \text{num_face}) \times 100 \quad (30)$$

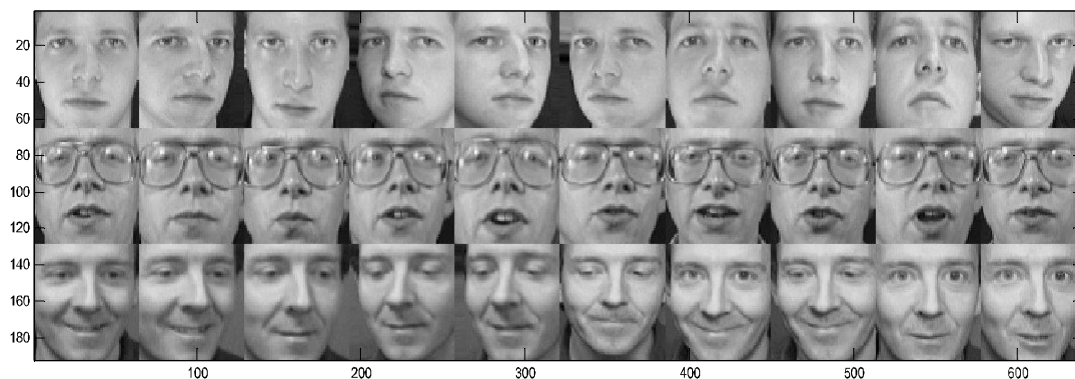
در رابطه بالا، num_face تعداد تصاویر چهره مورد آزمایش است که جهت شناسایی به سامانه داده می‌شود و num_correct تعداد چهره‌هایی است که به درستی شناسایی می‌شوند. در ابتدا به اختصار داده‌های چهره مورد استفاده را شرح می‌دهیم. تصاویر مورد استفاده در آزمایش‌ها، دارای ابعاد 32×32 هستند. برای شناسایی چهره، ابتدا زیرفضای ویژگی با استفاده از روش استخراج ویژگی (کاهش ویژگی) به دست آمده و داده‌های چهره به داخل آن زیرفضا تصویر می‌شوند و سپس از طبقه‌بند نزدیکترین همسایگی (NN) با فاصله اقلیدسی برای شناسایی چهره جدید (نمونه ازمون) استفاده می‌شود.

داده Yale حاوی ۱۶۵ تصویر با ۲۵۶ سطح خاکستری از چهره پانزده فرد مختلف است. از هر فرد یازده تصویر در شرایط مختلف نوری، حالات مختلف چهره، با عینک و بدون عینک اخذ شده است. داده ORL حاوی چهارصد تصویر از چهره چهل فرد است. از هر فرد ده تصویر در ژست و حالات مختلف چهره اخذ شده است. در شکل‌های (۲ و ۳)، به ترتیب نمونه‌هایی از تصاویر داده Yale و ORL نشان داده شده‌اند. جهت استخراج ویژگی و طبقه‌بندی (شناسایی) چهره‌ها از آن جایی که در داده چهره، هر نمونه یک تصویر است، در ابتدا هر تصویر را سطر به سطر یا ستون به ستون تبدیل به یک بردار ویژگی می‌کنیم. به عنوان مثال، یک تصویر 32×32 تبدیل به یک بردار ۱۰۲۴ بُعدی می‌شود.



(شکل-۲): تصاویر نمونه بانک داده Yale

(Figure-2): sample images of Yale dataset.



(شکل-۳): تصاویر نمونه بانک داده ORL

(Figure-3): sample images of ORL dataset.

می‌آورند. روش‌های NSFE، CSLDA، FSDA و DA-PC1 جزء این دسته هستند شکل (۴) دسته‌بندی روش‌های پیشنهادی را نشان می‌دهد.

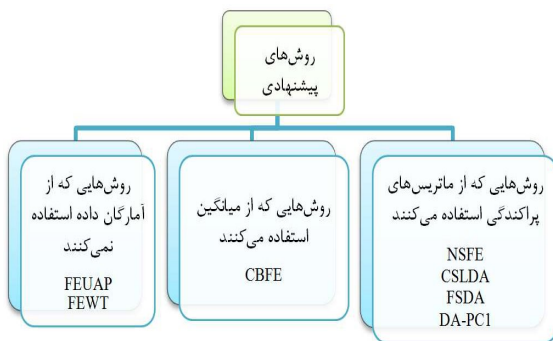
در شکل‌های (۵ تا ۸)، دقت شناسایی به‌دست آمده در برابر تعداد ویژگی‌های مختلف استخراج شده با استفاده از سه و پنج نمونه آموزشی به‌دست آمده‌اند. علاوه بر روش‌های LDA، NWFE، MMLDA و LPP، روش‌های استخراج ویژگی بر مبنای بیان داده و تبدیل مؤلفه‌های اصلی و نسخه‌های بهبود یافته آن‌ها نیز مورد بررسی و مقایسه با روش‌های مورد مطالعه قرار گرفته‌اند. نسخه غیرخطی (بر مبنای کرنل) روش PCA به نام KPCA^۱ در [22] معرفی شده است. پس از آن نسخه دیگری تحت عنوان PCA دوبعدی بر مبنای کرنل (K2DPCA)^۲ برای حل هر دو مشکل خطی بودن و بیان برداری PCA در [23] معرفی شده است. PCA یک روش بر مبنای بیان داده است که به اطلاعات تفکیک پذیری بین طبقه‌ها توجهی نمی‌کند. در [4] یک نسخه بهبودیافته از PCA پیشنهاد شده است که بر مبنای کمینه‌سازی

ما در این‌جا، هفت مورد از روش‌های پیشنهادی استخراج ویژگی برای داده ابرطیفی، NSFE، CBFE، FEUAP، FSDA، CSLDA، FEWT و DA-PC1 را با هم مقایسه کرده‌ایم. در روش FEUAP از معیار فاصله برای تعیین نقاط جذب استفاده شده است. هم‌چنین چهار روش استخراج ویژگی پرکاربرد LDA، NWFE، MMLDA و LPP را هم مورد آزمایش و مقایسه با روش‌های جدید قرار دادیم. آزمایش‌ها با استفاده از سه و پنج نمونه آموزشی در هر طبقه انجام شده است.

هفت روش پیشنهادی استخراج ویژگی را می‌توان در سه دسته کلی جای داد. دسته نخست، روش‌هایی هستند که برای استخراج ویژگی از داده، تنها از خود نمونه‌های آموزشی استفاده کرده و از آمارگان داده بهره‌ای نمی‌برند. روش‌های FEUAP و FEWT در این دسته قرار می‌گیرند. دسته دوم روش‌هایی هستند که برای استخراج ویژگی از میانگین (آمارگان مرتبه نخست) نمونه‌های آموزشی در هر طبقه بهره می‌برند. روش CBFE جزء این دسته محسوب می‌شود. دسته سوم، روش‌هایی هستند که برای استخراج ویژگی، تخمین‌های از ماتریس‌های پراکندگی (آمارگان مرتبه دوم) را به‌دست

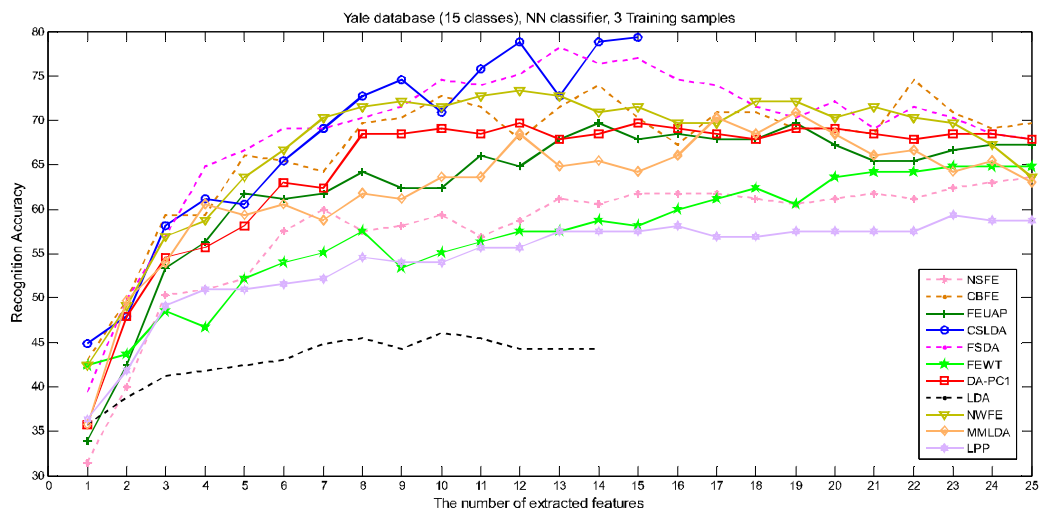
¹ Kernel based PCA (KPCA)

² kernel version of the 2DPCA method (K2DPCA)

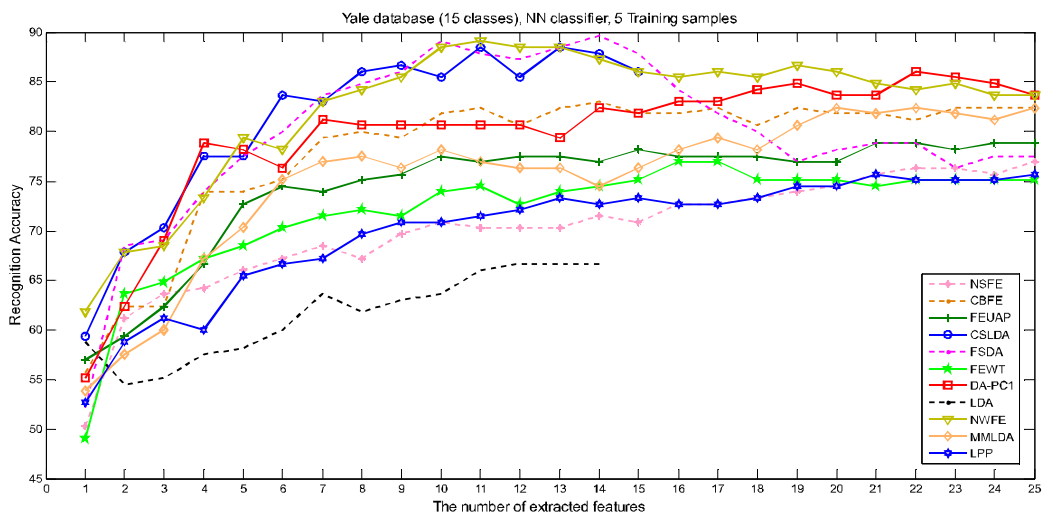


(شکل-۴): دسته‌بندی روش‌های پیشنهادی استخراج ویژگی
(Figure-4): Dividing the proposed feature extraction methods

پراکندگی‌های درون طبقه‌ای عمل می‌کند. این روش تحت عنوان PCA دوبعدی بر مبنای طبقه^۱ (C2DPCA) شناخته شد. مقایسه روش‌های مورد مطالعه با روش‌های پرکاربرد قبلی و همین‌طور روش‌های KPCA، K2DPCA و C2DPCA در جدول (۱) نمایش داده شده‌اند. بالاترین دقت شناسایی به‌دست‌آمده در میان ۲۵ ویژگی نخست برای هر دو داده با استفاده از سه و پنج نمونه آموزشی گزارش شده‌اند. اعداد داخل پرانتز، تعداد ویژگی‌هایی هستند که به‌ازای آن‌ها بالاترین دقت شناسایی به‌دست‌آمده است. مقایسه زمانی روش‌ها در شکل (۹) مشاهده می‌شود.

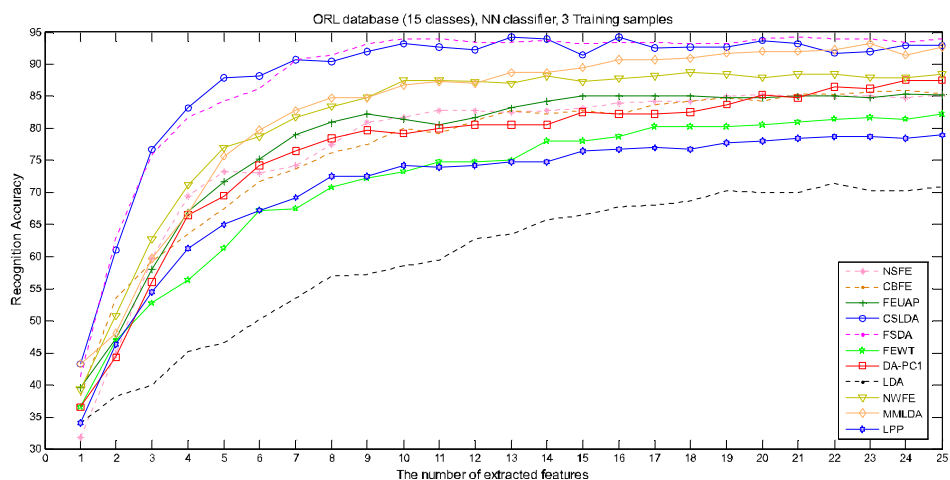


(شکل-۵): دقت شناسایی در برابر تعداد ویژگی‌ها با سه نمونه آموزشی برای داده Yale
(Figure-5): Recognition accuracy versus the number of features using 3 training samples for Yale dataset

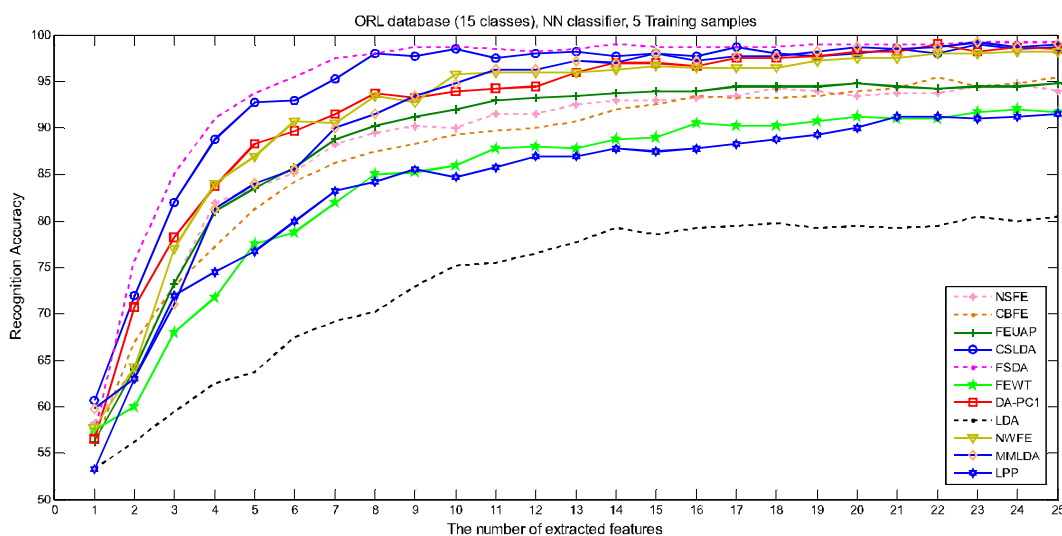


(شکل-۶): دقت شناسایی در برابر تعداد ویژگی‌ها با پنج نمونه آموزشی برای داده Yale
(Figure-6): Recognition accuracy versus the number of features using 5 training samples for Yale dataset

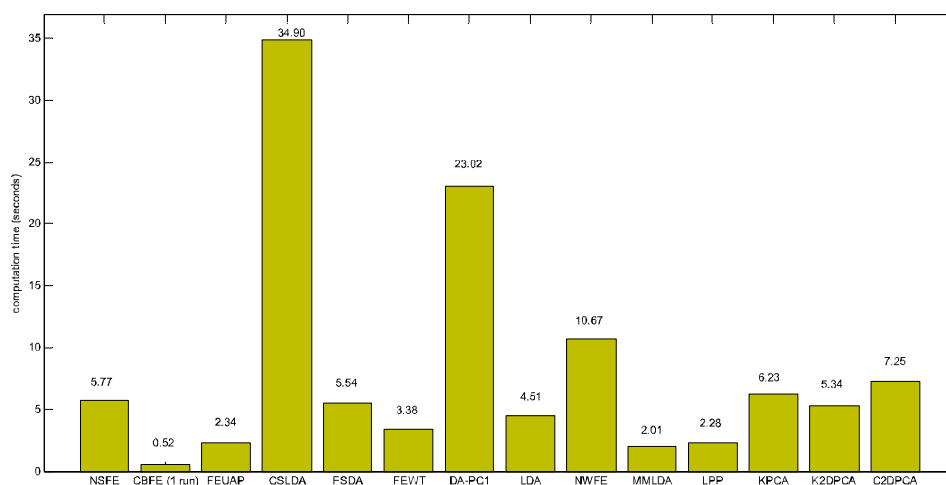
¹ Class-wise two-dimensional PCA (C2DPCA)



(شکل-۷): دقت شناسایی در برابر تعداد ویژگی‌ها با سه نمونه آموزشی برای داده ORL
(Figure-7): Recognition accuracy versus the number of features using 3 training samples for ORL dataset



(شکل-۸): دقت شناسایی در برابر تعداد ویژگی‌ها با پنج نمونه آموزشی برای داده ORL
(Figure-8): Recognition accuracy versus the number of features using 5 training samples for ORL dataset



(شکل-۹): زمان محاسبات در فرآیندهای استخراج ویژگی برای استخراج شش ویژگی از داده Yale با استفاده از سه نمونه آموزشی
(Figure-9): Computation time in the feature extraction processes for extraction of 6 features using 3 training samples in Yale dataset

(جدول-۱): بالاترین دقت شناسایی (در میان ۲۵ ویژگی نخست) به‌دست‌آمده توسط روش‌های مختلف

(Table-1): The highest recognition accuracy (among the first 25 features) achieved by different methods

Data	No. of Training	NSFE	CBFE	FEU AP	CSL DA	FSD A	FEW T	DA-PC1	LDA	NWF E	MM LDA	LPP	KPC A	K2D PCA	C2D PCA
Yale	3	63.64 (25)	74.55 (22)	69.70 (14)	79.39 (15)	78.18 (13)	64.85 (23)	69.70 (12)	46.06 (10)	73.33 (12)	70.91 (19)	59.39 (23)	72.13 (8)	75.54 (7)	78.87 (6)
	5	76.97 (25)	83.03 (14)	78.79 (21)	88.48 (11)	89.70 (14)	76.97 (16)	86.06 (22)	66.67 (12)	89.09 (11)	82.42 (20)	75.76 (21)	79.44 (10)	83.45 (9)	88.93 (9)
ORL	3	85.25 (20)	86.00 (24)	85.50 (24)	94.25 (13)	94.25 (21)	82.25 (25)	87.50 (24)	71.50 (22)	88.75 (18)	93.25 (23)	79.00 (25)	89.26 (13)	90.23 (11)	93.03 (10)
	5	94.75 (24)	95.50 (22)	94.75 (20)	99.00 (23)	99.25 (23)	92.00 (24)	99.00 (22)	80.50 (23)	98.25 (24)	99.25 (23)	91.50 (25)	93.22 (16)	94.45 (12)	97.54 (11)

کارایی کم LPP به‌دلیل این است که LPP از اطلاعات جدایی‌پذیری طبقه‌ها برای تمیز بین چهره‌ها استفاده نمی‌کند و تنها سعی در حفظ ساختار داده در فضای ویژگی بُعد کاهش‌یافته دارد.

مقایسه زمانی بین روش‌ها نشان می‌دهد که به‌ترتیب DA-PC1، CSLDA و NWFE کمترین روش‌ها هستند. توجه کنید که زمان درج‌شده برای روش CBFE بازای یک بار اجرای الگوریتم kmeans است. با نگاهی اجمالی به نتایج دقت و سرعت در می‌یابیم که درکل، روش‌های دقیق‌تر، به زمان محاسباتی بیشتری نیاز دارند.

۴- نتیجه‌گیری

در این مقاله، برای نخستین‌بار هفت روش استخراج ویژگی پیشنهادی برای داده‌های ابرطیفی سنجنش از دور را بر روی داده‌های چهره آزمایش کردیم. در میان این روش‌ها، CSLDA و FSDA در بیشتر موارد، دقت شناسایی بیشتری فراهم می‌کنند. از نظر زمانی، CSLDA از میان همه روش‌ها به زمان محاسباتی بیشتری نیاز دارد. روش FSDA که از دو معیار افزایش پراکندگی بین ویژگی‌ها (کاهش اطلاعات زائد و هم‌پوشان بین ویژگی‌های استخراجی) و افزایش جدایی‌پذیری طبقه‌ها (بیشینه‌کردن پراکندگی‌های بین‌طبقه‌ای و کمینه‌کردن پراکندگی‌های درون‌طبقه‌ای) به‌طور هم‌زمان بهره می‌برد، در کنار کارایی خوب، دارای زمان محاسباتی معقولی است. سایر روش‌های پیشنهادی، CBFE، DA-PC1، FEUAP، NSFE و FEWT هستند که همگی نسبت به روش‌های LPP نظارت‌شده و LDA دارای کارایی بهتری هستند و دو روش CBFE و DA-PC1 گاهی از دو روش قدرتمند NWFE و MMLDA نیز جلو می‌زنند.

نتایج آزمایش‌ها نشان می‌دهند به‌طورکلی از میان روش‌های پیشنهادی، FSDA و CSLDA و سپس DA-PC1 و CBFE کارایی خوبی دارند. روش‌های FSDA، CSLDA و DA-PC1 از آمارگان مرتبه دوم (ماتریس کواریانس) داده و روش CBFE از آمارگان مرتبه نخست داده (بردار میانگین) بهره می‌برند. از این‌رو، به‌نظر می‌رسد که اطلاعات با ارزشی از داده چهره در آمارگان مرتبه دوم (ماتریس‌های پراکندگی یا کواریانس) آن‌ها وجود دارد. در میان روش‌های پیشنهادی، FEUAP و FEWT که از آمارگان داده استفاده نمی‌کنند، نسبت به سایر روش‌های پیشنهادی ما دارای دقت کمتری هستند. در نتیجه، مشاهده می‌شود، روش‌هایی که از آمارگان داده بهره می‌برند، کارایی بهتری نسبت به سایر روش‌هایی دارند که از این اطلاعات استفاده نمی‌کنند. البته روش NSFE که از تخمین ماتریس‌های پراکندگی (آمارگان مرتبه دوم) داده استفاده می‌کند، نسبت به سایر روش‌هایی که ماتریس‌های پراکندگی را تخمین می‌زنند، CSLDA، FSDA و DA-PC1 کارایی کمتری دارد. برتری روش CSLDA به‌دلیل استفاده از توانایی نمونه‌های آزمایشی در کنار نمونه‌های آموزشی، برتری FSDA به‌دلیل بیشینه‌کردن پراکندگی‌های بین ویژگی‌ها در کنار بیشینه‌کردن جدایی‌پذیری طبقه‌ها و برتری DA-PC1 به‌دلیل استفاده از توانایی مؤلفه‌های اصلی در حفظ ساختار اصلی داده است. این در حالی است که NSFE تنها از فاصله اقلیدسی ساده بین نمونه‌ها برای وزن‌دهی ماتریس‌های پراکندگی بهره می‌برد.

در میان سایر روش‌های پرکاربرد مورد بررسی نیز، NWFE، MMLDA، KPCA، K2DPCA و C2DPCA نتایج خوبی دارند و LDA و LPP به‌ترتیب دارای بدترین کارایی هستند. کارایی کم LDA به‌دلیل منفردشدن ماتریس پراکندگی درون‌طبقه‌ای با نمونه‌های آموزشی کم است.

- [10] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. San Diego, CA, USA: Academic, 1990.
- [11] B.C. Kuo, D.A. Landgrebe, "Nonparametric weighted feature extraction for classification", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 5, pp. 1096-1105, 2004.
- [12] J. Xu, J. Yang, Z. Gu, and N. Zhang, "Median-mean line based discriminant analysis", *Neurocomputing*, vol. 123, pp. 233-246, 2014.
- [13] X.F. He, P. Niyogi, "Locality preserving projections", In: *Advances in Neural Information Processing System*, vol. 16, pp. 153-160, 2004.
- [۱۴] ایمانی، مریم، قاسمیان، حسن، "استخراج ویژگی نظارت شده غیرپارامتریک برای طبقه‌بندی تصاویر ابرطیفی با نمونه آموزشی محدود"، فصل‌نامه صنایع الکترونیک، دوره ۴، شماره ۳، پاییز ۱۳۹۲.
- [14] M. Imani, H. Ghassemian, "Nonparametric Supervised Feature Extraction for Classification of Hyperspectral Images Using Limited Training Samples", *Electronics Industries Quarterly*, vol. 4, no. 3, Autumn 2013.
- [15] M. Imani, H. Ghassemian, "Band Clustering-Based Feature Extraction for Classification of Hyperspectral Images Using Limited Training Samples", *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 8, pp. 1325-1329, 2014.
- [16] M. Imani, H. Ghassemian, "Feature Extraction Using Attraction Points for Classification of Hyperspectral Images in a Small Sample Size Situation", *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 11, pp. 1986-1990, 2014.
- [۱۷] ایمانی، مریم، قاسمیان، حسن، "طبقه‌بندی تصاویر ابرطیفی با استفاده از تحلیل ممیز خطی فضای خوشه و مجموعه نمونه‌های آموزشی کوچک"، مجله نشریه مهندسی برق و مهندسی کامپیوتر ایران، سال ۱۴، شماره ۱، صفحات ۸۱-۷۳، بهار ۱۳۹۵.
- [17] M. Imani, H. Ghassemian, "Classification of Hyperspectral Images Using Cluster Space Linear Discriminant Analysis and Small Training Set", *Iranian Journal of Electrical and Computer Engineering*, vol. 14, no. 1, pp. 73-81, June 2016.
- [18] M. Imani, H. Ghassemian, "Feature space discriminant analysis for hyperspectral data feature reduction", *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 102, pp. 1-13, 2015.
- [19] M. Imani, H. Ghassemian, "Feature Extraction Using Weighted Training Samples", *IEEE*
- [1] P. Huang, C. Chen, Z. Tang, and Z. Yang, "Discriminant similarity and variance preserving projection for feature extraction", *Neurocomputing*, vol. 139, pp. 180-188, 2014.
- [2] S. Tan, X. Sun, W. Chan, L. Qu, and L. Shao, "Robust Face Recognition With Kernelized Locality-Sensitive Group Sparsity Representation", *IEEE Transactions on Image Processing*, vol. 26, no. 10, pp. 4661-4668, Oct. 2017.
- [3] Y. Shen, M. Yang, B. Wei, C. T. Chou and W. Hu, "Learn to Recognise: Exploring Priors of Sparse Face Recognition on Smartphones", *IEEE Transactions on Mobile Computing*, vol. 16, no. 6, pp. 1705-1717, June 2017.
- [4] C. Guzel Turhan and H. S. Bilge, "Class-wise two-dimensional PCA method for face recognition", *IET Computer Vision*, vol. 11, no. 4, pp. 286-300, 2017.
- [5] W. Wang, R. Wang, Z. Huang, S. Shan and X. Chen, "Discriminant Analysis on Riemannian Manifold of Gaussian Distributions for Face Recognition With Image Sets", *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 151-163, Jan. 2018.
- [6] Yang, W.-H., Dai, D.-Q., "Two-Dimensional Maximum Margin Feature Extraction for Face Recognition", *IEEE Transactions on Systems, Man, and Cybernetics—Part B: Cybernetics*, vol. 39, no. 4, pp. 1002-1012, 2009.
- [۷] احمدخانی سمیه، ادیبی پیمان، "مدل ترکیبی تحلیل مؤلفه اصلی احتمالاتی بانظارت در چارچوب کاهش بعد بدون اتلاف برای شناسایی چهره"، فصل‌نامه پردازش علائم و داده‌ها، دوره ۱۲، شماره ۴، صفحات ۶۵-۵۳، ۱۳۹۴.
- [7] S. Ahmadkhani, P. Adibi, "Supervised Probabilistic Principal Component Analysis Mixture Model in a Lossless Dimensionality Reduction Framework for Face Recognition", *Quarterly Journal of Signal and Data Processing*, vol. 12, no. 4, pp. 53-65, 2016.
- [8] Yang, M., Zhang, L., Shiu, S. C.-K., and Zhang, D., "Monogenic Binary Coding: An Efficient Local Feature Extraction Approach to Face Recognition", *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 6, pp. 1738-1751, 2012.
- [9] G. F. Hughes, "On the mean accuracy of statistical pattern recognition," *IEEE Transactions on Information Theory*, vol. IT-14, no. 1, pp. 55-63, 1968.

Geoscience and Remote Sensing Letters, vol. 12, no. 7, pp. 1387 – 1386, 2015.

- [20] M. Imani, H. Ghassemian, "Feature reduction of hyperspectral images: discriminant analysis and the first principal component", *Journal of AI and Data Mining*, vol. 3, no. 1, pp.1-9, 2015.
- [21] G. H. Golub, and C. F.van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD, USA: The Johns Hopkins Univ. Press, 1996.
- [22] M. Yang, N. Ahuja, and D. Kriegman, "Face recognition using kernel eigenfaces", *Proc. International Conference on Image processing*, 2000, pp. 37–40.
- [23] V. D. M Nhat, and S. Lee, "Kernel-based 2DPCA for Face Recognition", *Proc. IEEE International Symposium on Signal Processing and Information Technology*, December. 2007, pp. 35–39.



مریم ایمانی کارشناسی و کارشناسی

ارشد خود را در رشته مهندسی برق-گرایش مخابرات از دانشگاه شاهد دریافت کرد. وی در سال ۱۳۹۴ و ۱۳۹۵ به ترتیب دوره دکترا و پسادکترای خود را در دانشگاه

تربیت مدرس به پایان رساند. ایشان از سال ۱۳۹۶ استادیار دانشکده مهندسی برق و کامپیوتر دانشگاه تربیت مدرس هستند. زمینه‌های پژوهشی مورد علاقه ایشان، شناسایی آماری الگو، پردازش سیگنال و اطلاعات و مهندسی سنجش از دور است.

نشانی رایانامه ایشان عبارت است از:

maryam.imani@modares.ac.ir



محمد حسن قاسمیان یزدی تحصیلات

کارشناسی خود را در رشته مهندسی مخابرات در سال ۱۳۵۸ در دانشکده مخابرات تهران به پایان رساند. ایشان مدرک کارشناسی ارشد و دکترای خود را

در رشته مهندسی مخابرات در دانشگاه پردو آمریکا به ترتیب در سال ۱۳۶۳ و ۱۳۶۷ اخذ کرد. ایشان هم‌اکنون استاد دانشکده مهندسی برق و کامپیوتر هستند. آنالیز و پردازش تصاویر از منابع چندگانه، پردازش اطلاعات و شناسایی الگو، مهندسی سنجش از دور، پردازش تصاویر و سیگنال‌های مهندسی پزشکی از جمله علاقه‌های پژوهشی ایشان است.

نشانی رایانامه ایشان عبارت است از:

ghassemi@modares.ac.ir

